



**UC** Leuven  
Limburg  
**MOVING MINDS**

## Levels of aggregation: Identification of sociolinguistic determinants of written L1 proficiency in higher education

Jose Tummers<sup>1,2</sup>, Annelies Deveneyns<sup>1</sup>

<sup>1</sup> UC Leuven-Limburg

<sup>2</sup> KU Leuven



## Contents

1. Problem statement
2. Data gathering
3. Written language proficiency in HE
4. Sociolinguistic determinants of written language proficiency in HE
5. Discussion



## 1. Problem statement

- **Research on language proficiency**
  - Focus on foreign languages
  - Mother tongue
    - Special targets groups (cf. community integration)
    - Language development (up to end secondary school)
- **Language proficiency teaching in Flemish education**
  - Secondary education: mother tongue, foreign languages
  - Higher education
    - Foreign languages: focus on domain-specific/professional competences
    - Mother tongue: rare; mainly conceived as professional communication skills



## 1. Problem statement

- **Wider context of higher education in Flanders**
  - **Research into L1 proficiency in HE:** lack of L1 writing proficiency (Berckmoes & Rombouts 2009; Deveneyns & Tummers 2013; Peeters & Van Houtven 2010; Van Houtven et al. 2010)
  - **Transition SE to HE:** gap between final qualifications SE and implicit prerequisites HE (Bogaert & Verheyden 2011; Van den Branden 2010)
  - Problem = **academic language use**
- Importance of mother tongue proficiency in HE: correlation between L1 language proficiency and study progress (De Wachter & Heeren 2013)



## 1. Problem statement

- **Research questions**

- *What is the level of the written L1 proficiency achieved by first year students in HE in Flanders?*
- *What are the sociolinguistic determinants of written L1 proficiency? Are they characteristic of the individual students or of does it concern group properties?*



## 2. Data gathering

- Corpus of texts written by **first year students** at the end of 1<sup>st</sup> BA
- University College Leuven
  - University College/University of Applied Sciences: student population is a better representation of SE graduates, compared to that of universities (where, in general, only the best students start)
  - 13 study programs in domains of Business Management, Teacher Training, Social Work, Health Care and Technology
- **Cluster sampling** (McDaniel & Gates 2007)



## 2. Data gathering

- **Writing task**
  - Reflection of settings of student text writing (ecological validity)
    - Text rather than multiple choice exercise, fill-in exercises (Hyland 2001; Mukherjee 2006)
    - Computer, all sources deemed necessary
  - Argumentative text: "Convince your audience of your opinion on ..."
  - Familiar subject: social network sites
  - Length: 500w; Duration: max. 1h
- **Resulting dataset: 346 texts**
  - Subject familiarity (on 5-point Likert scale:  $M = 3.6$ ,  $m = 4$ )
  - Time (on 5-point Likert scale:  $M = 3.7$ ,  $m = 4$ )



## 2. Data gathering

- **Evaluation written language proficiency: analytic evaluation**
  - 4 criteria present in assignment
    - Language errors
    - Text structure, textual organization
    - Argumentation
    - Persuasion
  - 4-point scales:  $-2 < -1 < +1 < +2$ 
    - Definition of features on every level (Knoch 2001)
    - Score '+1'  $\Leftrightarrow$  level B1 of CEFL (CNaTV n.d.; European Commission 2001)



## 2. Data gathering

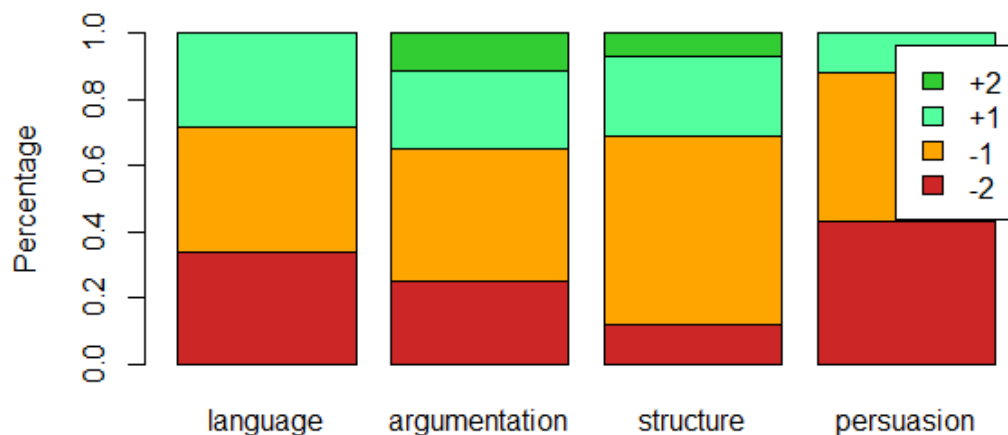
- **Sociolinguistic information student**

- Gender: male vs. female
- Age secondary degree
- Secondary education diploma: gse (general SE), tse (technical SE), pse (professional SE)
- Home language: Dutch, no Dutch
- Level diploma mother: SE, HE
- Background HE: generation student, university, university of applied sciences



## 3. Written language proficiency

- **Overview of scores per analytic criterion**





### 3. Written language proficiency

- **Integrated score – Combining the 4 analytical criteria**
  - Reduction to holistic score: loss of information of analytical criteria
  - Solution: integrated score computed by Multiple Correspondence Analysis (PCA for categorical data) (Greenacre 2003, 2006) – `ca` library in R (Nenadic & Greenacre 2007)



### 3. Written language proficiency

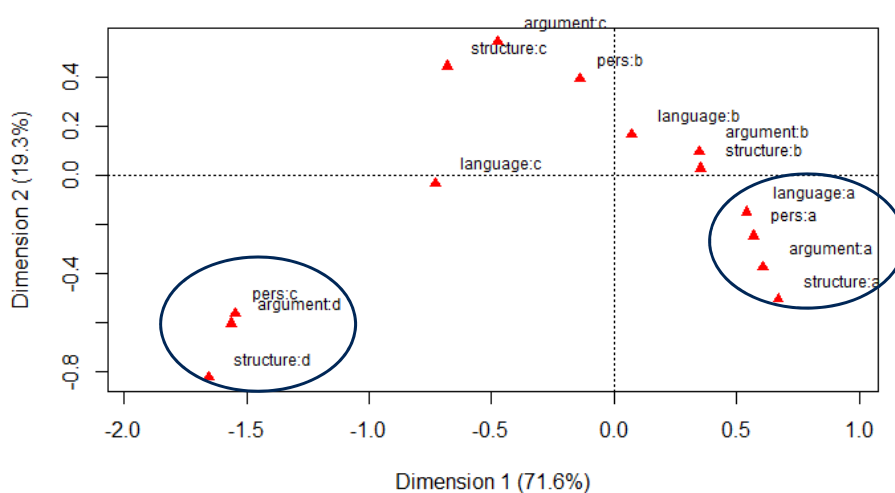
- **Integrated score – Summary MCA**

Dimension	Value	%	Cum %	Scree plot
1	0.398203	71.6	71.6	*****
2	0.107269	19.3	90.8	*****
3	0.016881	3.0	93.9	*
<b>Total</b>	<b>0.556401</b>			



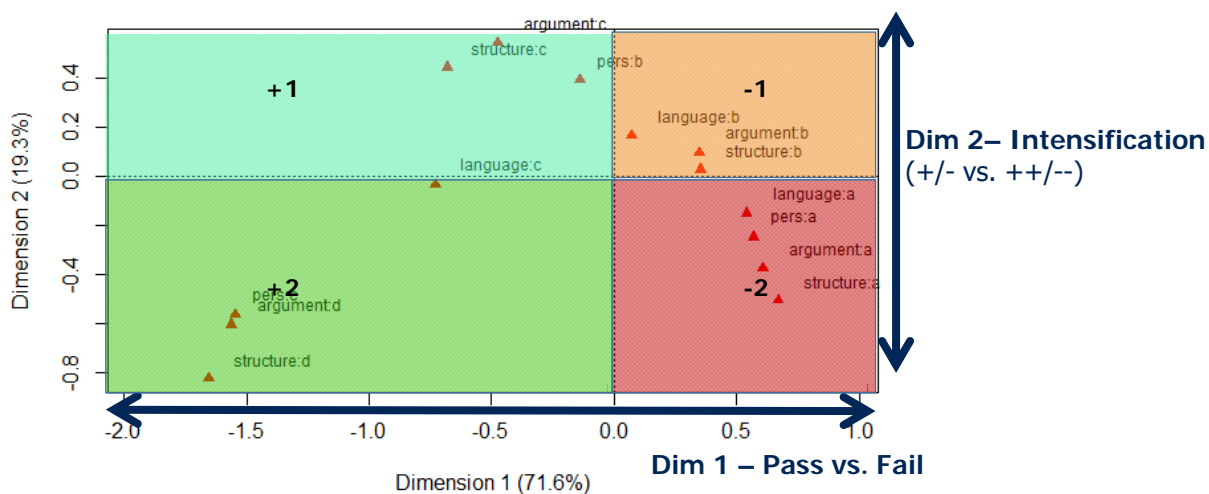
### 3. Written language proficiency

- Integrated score – MCA plot (first 2 dimensions)



### 3. Written language proficiency

- Integrated score – MCA plot (first 2 dimensions)





### 3. Written language proficiency

- **Integrated score – Summary**
  - **Overall result ≠ satisfactory**: majority of the students is unable to write a satisfactory argumentative text
  - **High correlation between grades on 4 criteria** ⇒ upon entering HE, youngsters seem to have reached a stable level of language proficiency across all/most criteria
- Analysis of sociolinguistic determinants: based on the scores on dimension 1 (explaining 71,6% of the inertia)



### 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

- **Model specification**
  - Response variable: score on Dim 1 of MCA
  - Explanatory variables
    - Individual **student properties (level-1)**: gender, secondary education diploma, age secondary education diploma, language at home, background HE, diploma mother
    - **Group properties (level-2)**: study program HE, school SE
  - Hierarchical model (Baayen 2008; Gelman & Hill 2007)
  - Analysis: lme4 (Bates 2005) and arm (Gelman & Hill 2007) libraries in R





## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

- Results – student properties (level-1)

	Chisq	Df	Pr (>Chisq)
home.lan	7.5124	2	0.023372 *
Hei	1.5133	3	0.679199
edu.moth	1.6845	2	0.430739
gender	0.0037	1	0.951353
sec.edu	14.2975	3	0.002527 **
age.sec.edu	0.1369	1	0.711398



## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

- Results – student properties (level-1)

	Chisq	Df	Pr (>Chisq)
home.lan	7.5124	2	0.023372 *
hei	1.5133	3	0.679199
edu.moth	1.6845	2	0.430739
gender	0.0037	1	0.951353
sec.edu	14.2975	3	0.002527 **
age.sec.edu	0.1369	1	0.711398

Home language  
Dutch > not.Dutch

SE diploma  
GSE > TSE > PSE



## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

### • Results – group properties (level-2)

#### ▪ SE school

- No significant impact (comparison  $GLMM_{\text{ranef.StudyProgram}}$  &  $GLMM_{\text{ranef.studyprogram.SESchool}}$ :  $\Delta X^2 = 0$ ,  $df = 1$ ,  $p = 1$ )

#### ▪ HE study program

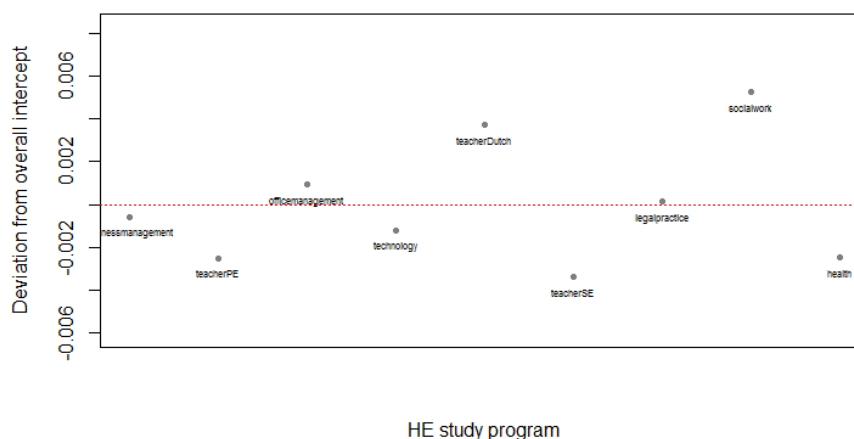
- Significant impact (comparison  $GLMM_{\text{ranef.StudyProgram}}$  & GLM:  $\Delta X^2 = 2033.497$ ,  $df = 78$ ,  $p < 0.00001$ )
- ICC = 0.082 (proportion of variance explained by HE study program)



## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

### • Results – group properties (level-2)

Random intercepts for HE study programs





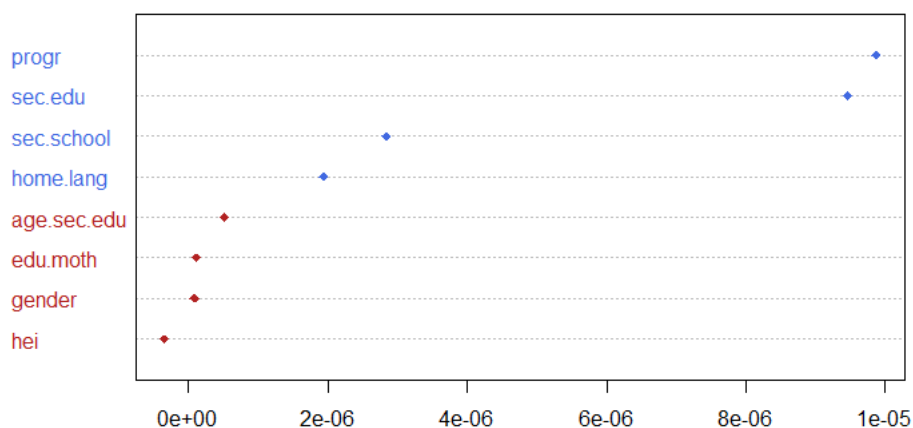
## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

- **Results – importance of variables (student and group properties taken together)**
  - Random forest analysis (Stroebel et al. 2008)
  - Ascertainment of relative impact of student-related variables (level-1) and grouping variables (level-2)
  - Output: plot with importance of different variables



## 4. Sociolinguistic constraints L<sub>1</sub> adult writing proficiency

- **Results – importance of variables (student and group properties taken together)**





## 5. Discussion

### Written L1 proficiency in Flemish HE

- **Negative overall image:** in general 1<sup>st</sup> BA university college students do not reach the standards set with respect to L1 proficiency
- **Necessity of L1 support (at start) in HE,** especially with respect to genre-specific characteristics of academic prose



## 5. Discussion

### Sociolinguistic constraints of written L1 proficiency in Flemish HE

- **Student-related determinants**
  - Secondary education diploma
  - Language situation at home
- **Group-related determinants**
  - Study program HE – linked to ECTS credits L1 in curriculum ( $r = 0.75$ )
  - School in SE
    - 👉 Opposite results for secondary school in hierarchical model and random forest analysis: a lot of SE schools with few respondents in data matrix (respondents per school:  $M = 3.46$ ,  $m = 1.5$ ,  $SD = 5.12$ )



## 5. Discussion

### Implication for language policy HEIs

- Integration of L1 language proficiency in 1<sup>st</sup> BA curriculum
  - Passive skills
  - Active skills
- Further research to identify specific subgroups of entering students with respect to L1 needs → tailored remediation trajectories



## 5. Discussion

### Further research

- Inclusion of other text genres
- Extension of the sample with students from other university colleges/universities of applied sciences and of universities
- Link of analytical assessment to linguistic features (Crosley & McNamara 2011)



## Further information

- UC Leuven-Limburg  
<http://www.ucll.be>
- KU Leuven, RU Quantitative Lexicology and Variational Linguistics  
<http://wwwling.arts.kuleuven.be/qlvl>
- Mail
  - [Jose.Tummers@ucll.be](mailto:Jose.Tummers@ucll.be)
  - [Annelies.Deveneyns@ucll.be](mailto:Annelies.Deveneyns@ucll.be)